Supporting Information

Sethi et al. 10.1073/pnas.0810961106

SI Text

Methods

Modeling. ***Bacterial pretransfer and posttransfer Glu complexes.*** The pretransfer state is based on the crystal structure of the adenylate (GluRS:tRNA^{Glu}:Glu-AMP) complex (PDB ID 1N78) (1), which contains the *T. thermophilus* GluRS, transcribed tRNA^{Glu}, and the adenylate analog glutamol-AMP. For the pretransfer complex, we used the program psfgen in VMD (2) to modify glutamol-AMP to Glu-AMP through the addition of a carbonyl oxygen. Modeling of pretransfer glutamyl-AMP from its analog is straightforward and produces minor changes in the structure of the active site. In the reaction mechanism proposed for GlnRS (3), the phosphate group on AMP acts as a general base and accepts a proton from the 2′ AMP OH of A76. The 2′ O of A76 forms an ester bond with the carbonyl group in the backbone of Gln to form the charged tRNA molecule along with the concomitant formation of HAMP. The proton may later get transferred to the surrounding water molecules in the pocket or to a conserved nearby histidine in the HIGH motif of the class I aaRSs. For the posttransfer complex GluRS:Glu-tRNA^{Glu}:AMP, the backbone carbonyl carbon was detached from the AMP and reconnected to the 2′ oxygen in the ribose of A76 to create Glu-tRNA^{Glu}. The minor side-chain rearrangements required to accommodate these ligand changes occurred during minimization. The posttransfer complex GluRS:Glu-tRNA^{Glu}:HAMP with protonated AMP was modeled similarly except that the AMP had a hydrogen added to the phosphate. During minimization of the GluRS:Glu-tRNA^{Glu}:HAMP complex, His 15 from the HIGH motif in GluRS forms a contact with HAMP, and this contact does not break during the 20-ns MD simulation.

Archaeal pretransfer (LeuRS:tRNA^{Leu}:Leu-AMP) complex. The structure of the archaeal LeuRS:tRNA^{Leu} complex from *P. horikoshii* in which the CCA-hairpin of the tRNA is in the active site of the catalytic domain has been resolved (PDB ID 1WZ2) (4) with the exception of a 16-aa loop in the editing domain from residues 357 to 372. We used Modeller (5) with loop optimization to model this loop in the LeuRS structure such that it does not clash with the tRNA. The Leu-AMP-binding site is conserved in all LeuRSs in structure and sequence. The structure of the bacterial LeuRS:tRNA^{Leu} complex with the sulfamoyl analog of Leu-AMP analog (PDB ID: 2V0C) (6) was used to place the Leu-AMP adenylate in the active site of the archaeal structure. Modeller was used to model the conformation of the side chain of the amino acids in the archaeal LeuRS structure interacting with Leu-AMP using the bacterial structure as a template.

Molecular Dynamics. ***System setups.*** The MD simulations of the solvated complexes were performed using NAMD2 (7) with the CHARMM27 force field (8). The histidine protonation states were predicted using the WHATIF server and visually checked (9). Parameters for the 2 aminoacyl-adenylate molecules were developed by analogy with parameters for the separate AMP and corresponding amino acid already present in the CHARMM27 force field. Parameters for the bond between Glu and A76 were developed by analogy with those for the Sep-A76 linkage (10). The parameters for HAMP were developed by analogy to the CHARMM27 parameters available for AMP and protonated pyrophosphate. The protein:RNA complexes were explicitly solvated with TIP3 water molecules (11) and ions added according to the protocol developed for molecular dynamics simulations of the EF-Tu:tRNA complex (12), which is summarized below.

Psfgen was used to add hydrogen atoms to the macromolecules. The systems were neutralized by placing Mg^{2+} and K^+ with the program ionize (www.ks.uiuc.edu/Development/MDTools/ionize), which places the ions at the minima of the Coulombic electrostatic interaction energy calculated on a uniform grid. The Mg^{2+} placement protocol developed previously placed 3 Mg^{2+} ions on the primary solvation shell of the tRNA (at a distance of 2 Å from the tRNA) (12). In addition, 14 Mg^{2+} and 44 K^+ ions were placed 6.5 Å from the tRNA in the GluRS simulations, whereas 17 Mg^{2+} and 64 K^+ ions were placed in the LeuRS simulation for the 88-mer tRNA^{Leu}. We placed water molecules near the Mg^{2+} ions to complete their primary solvation shells. The systems were solvated in a 2-step process. In the first step, Solvate 1.0 [Grubmueller H (1996) SOLVATE 1.0. www.mpibpc.gwdg.de/abteilungen/071/solvate/docu.html] was used with 2 Gaussians to add 2 layers of water molecules to the system. This solvated the inner catalytic pocket and protein:RNA docking interface. Then, the Solvate 1.2 plugin to VMD (2) was used to place the bulk water, resulting a full system size of $116 \times 80 \times 119$ Å^3 and ≈100,000 atoms in the GluRS simulations, whereas the system was $115 \times 118 \times 166$ Å^3 and 211,268 atoms in the LeuRS simulation.

Simulation protocol. All simulations were done with periodic boundary conditions using the NPT ensemble with pressure set to 1 atmosphere using the Langevin piston and temperature set to 298 K using Langevin dynamics. Electrostatics were calculated with the particle mesh Ewald method (13). The van der Waals interactions were calculated using a switching distance of 10 Å and a cutoff of 12 Å. Time steps for updates of bonded, van der Waals, and electrostatic calculations were 1, 2, and 4 fs, respectively.

Both the GluRS and LeuRS systems were minimized using a 4-step protocol in which the water molecules were allowed to associate with the macromolecule before allowing the macromolecule to move. These steps were: heavy atoms fixed (2,000 steps), heavy atoms fixed excluding water and ions (3,000 steps), macromolecule backbone atoms fixed (5,000 steps), and all atoms free to move (20,000 steps). During the initial equilibration, the system was gradually heated to 298 K (12, 14) during which different parts of the system were harmonically constrained. The initial temperature was set to 100 K, and ions and heavy atoms in the protein and nucleic acid chains were harmonically constrained for the first 25,000 fs. Then the temperature was raised to 200 K, and backbone atoms were harmonically constrained for 25,000 fs. Force constants for all harmonic constraints were set to 1 kcal mol^{-1}Å^2. Finally, the temperature was raised to 298 K, and all atoms were freed for the next 3.9 ns. After this 4-ns equilibration, each system was run for 16 ns.

Correlation and Principal Component Analyses. Correlations between all of the residues and nucleotides in the aaRS:tRNA complex were analyzed for the last 16 ns of the 20-ns MD trajectory using the normalized covariance:

$$C_{ij} = \frac{Cov_{ij}}{(\langle \Delta\vec{r}_i(t)^2 \rangle \langle \Delta\vec{r}_j(t)^2 \rangle)^{1/2}}$$

where $Cov_{ij} = \langle \Delta\vec{r}_i(t) \cdot \Delta\vec{r}_j(t) \rangle$ and $\Delta\vec{r}_i(t) = \langle \vec{r}_i(t) - \vec{r}_i(t) \rangle$. $\vec{r}_i(t)$ is the position vector of the C_α or P atom of the ith residue or nucleotide of the protein or tRNA, respectively, at time t, and $\langle . \rangle$ refers to the time average of the quantity within the brackets.

This correlation matrix is also called the dynamic cross-correlation matrix, which has been used to characterize the correlation in motion of protein residues (15–18). The correlations between the residues fall in the range from −1 to 1. If the residues move in the same direction in most of the frames, the motion is considered to be correlated, and C_{ij} will be positive. If they move in opposite directions in most frames, the motion is said to be anticorrelated, and C_{ij} will be negative. If the correlation value between the 2 residues is close to zero, then the motion is said to be uncorrelated. If the 2 residues move in perpendicular directions in a coordinated fashion, their correlation value will also be close to zero, which is an artifact of this correlation function. The frames are saved at an interval of every 0.5 ps, and a total of 32,000 frames was analyzed for the correlation matrices of each simulation.

To investigate the collective behavior within the complex, a standard principal component analysis (PCA) of the motions of the C_α and the P atoms during the equilibration was performed as implemented in the program CARMA (19). The unnormalized covariance matrix, **Cov** defined above, is obtained by averaging over the last 16 ns of the 20-ns MD simulation. Diagonalization of the covariance matrix provides the largest eigenvalues and their accompanying eigenvectors, which capture the largest fraction of the observed variance in the motion. The contribution of each eigenvector to the overall motion is obtained from the projection matrix (Figs. S3 and S4).

On projecting the data from principal component i onto the Cartesian coordinates, the RMSD of each residue was calculated due to the ith principal component. The RMSD plots give an estimate of regions that are highly coupled due to the ith principal component (Figs. S3 and S4).

Characteristic Path Length (CPL) Analysis. The CPL is defined as the average distance between all pairs of nodes in the network.

$$CPL = \frac{\Sigma_{i,j} D^0{}_{(i,j)}}{N_{\text{pairs}}}$$

where N_{pairs} is the number of all pairs of nodes connected by a path in the system. The contribution of a node k to communication within the network or at the interface is found by recalculating the CPL after removing node k or its contacts at the interface (protein:RNA contacts) ($CPL_{k,\text{rem}}$) from the network, respectively. The difference in CPL ($\Delta CPL_{k,\text{rem}} = CPL_{k,\text{rem}} - CPL$) had previously been used to predict residues important in allosteric signaling within protein families (20). A Z score analysis provides a measure for the relative change in CPL:

$$Z_k = \frac{\Delta CPL_{k,\text{rem}} - \langle \Delta CPL_{k,\text{rem}} \rangle}{\sigma_{\Delta CPL_{k,\text{rem}}}}$$

where $\langle \Delta CPL_{k,\text{rem}} \rangle$ and $\sigma_{\Delta CPL_{k,\text{rem}}}$ refer to the average and standard deviation of $\Delta CPL_{k,\text{rem}}$ across all of the nucleotides in the tRNA or all of the residues in the protein, respectively.

Community Repartition Difference. For the GluRS:tRNA interface mutational analysis as well as examination of the robustness of the community algorithm, a community repartition difference was used to quantify the difference between a reference community partition (c_1) and a new partition (c_2) of the same set of nodes. This partition measure is similar to that developed by Knox (21) and Rand (22). The community repartition difference (CRD) is defined as

$$\text{CRD}(c_1, c_2) = 1 - \frac{\sum_{n_i,n_j} z(n_i, n_j, c_1) z(n_i, n_j, c_2)}{\sum_{n_i,n_j} z(n_i, n_j, c_1)},$$

where $z(n_i, n_j, c_1)$ is 1 if nodes n_i and n_j are in the same community in community network c_1 and 0 otherwise. This is a normalized count of node pairs that are grouped together in both community networks. If the community partitions are exactly the same, the CRD will be 0, and if they are totally different, the CRD will be 1.

Evolutionary Analysis. Evolutionary analyses of the structures and sequences of the catalytic domains for both Class I and Class II aaRSs (23, 24) have already been performed and are in good agreement with the phylogenetic analysis of the entire sequences (25). The evolutionary analysis is performed here to measure the conservation of the residues and nucleotides that are predicted to be important for allostery in this work. Because the ACB domain in the bacterial version of GluRS has a different fold from the archaeal and eukaryotic versions, an evolutionary profile (EP) was created only for the bacterial sequences. Similarly, as the identity elements for LeuRS differ among the 3 domains of life, the mechanism for allosteric signal transduction could also vary among the domains. Hence, an EP for the archaeal version of LeuRS was also created. We used the sequences of GluRS from the *T. thermophilus* and LeuRS sequence from the *P. horikoshii* to perform BLAST searches (26) over the NCBI nonredundant database (27), with E value cutoff of 10^{-5} to construct the respective EPs. The sequences were filtered for bacterial and archaeal domains of life in the GluRS and the LeuRS database search, respectively. ClustalW (28) was used to align each sequence set, and the alignments were manually improved. Residues R358 and G444 (*T. thermophilus* numbering) have been previously identified to differentiate between the D- and ND-GluRS sequences, respectively (29). These residues were used to split the GluRS sequences into the D-GluRS and the ND-GluRS in this work. The EP for D-GluRS (Table S8) was created by running SeqQR at a cutoff of 50% for the bacterial D-GluRS, which returned a set of 35 D-GluRS sequences. The EP for the archaeal LeuRS (Table S9) was created by running SeqQR at a cutoff of 75%, which returned a set of 26 LeuRS sequences. SeqQR selects a set of the most linearly independent set of sequences to form a statistically well-balanced set that represents the phylogenetic diversity with the minimal number of sequences (24). The percent conservation reported here is based on these representative sets.

The tRNA EPs were also created only from bacterial $tRNA^{Glu}$ and archaeal $tRNA^{Leu}$ sequences (Table S8 and Table S9) that included all of the different isoacceptors. The EPs were created using a similar methodology to that used in ref. 12. The MultiSeq (30) plugin in VMD (2) was used to perform all of the steps in the EP construction and analysis.

Additional Background

Identity Elements in $tRNA^{Glu}$. Many of the identity elements for the different tRNA species have been identified experimentally (31). For the $tRNA^{Glu}$ in *E. coli*, the experimentally determined identity elements are found to be base pairs 1·72, 2·71, and 4·69 in the acceptor stem along with bases 34–37 in the anticodon loop and the U11·A24 base pair in addition to the U13·G22·A46 triplet in the GG arm (32, 33). All of the identity elements are highly conserved in the $tRNA^{Glu}$s, whereas bases 35 and 36 in the anticodon are completely conserved. In most organisms there are 2 different $tRNA^{Glu}$ genes that differ in the wobble base position of the anticodon (U34 and C34). U34 is usually modified inside the cell to maintain its affinity to GluRS (34), whereas no such modification is known for C34. The structure of the GluRS:$tRNA^{Glu}$ complex with an adenylate analog (PDB ID: 1N78) from *Thermus thermophilus* has been resolved with the anticodon CUC for the transcribed tRNA species (35).

GluRS and LeuRS Phylogeny. The phylogenetic distribution of the GluRS and LeuRS catalytic domains are found to be canonical

in nature in which the 3 domains of life (*Archaea*, *Bacteria*, and *Eukarya*) form separate clusters with the eukaryal and archaeal versions of the molecule being closer to each other (25). In addition, the GlnRSs diverge from the eukaryal portion of the GluRS tree, indicating that they were a late innovation (10). The bacterial GluRS has an α-helical ACB domain (35), whereas the ACB domain in archaeal and eukaryal GluRSs is homologous to the β-sheet ACB domain of GlnRS (36). Some organisms aminoacylate $tRNA^{Gln}$ with glutamate and then convert the glutamate into glutamine on the tRNA using a second enzyme in the indirect route for aminoacylation (37). These GluRSs are called the nondiscriminate GluRSs (ND-GluRSs). All archaeal organisms and some bacterial phyla use the ND-GluRS to charge both $tRNA^{Glu}$ and $tRNA^{Gln}$ with glutamate. The GluRSs that can discriminate between $tRNA^{Glu}$ and $tRNA^{Gln}$ and only charge the cognate $tRNA^{Glu}$ with glutamate are called the discriminate GluRSs (D-GluRSs). Recently, it was discovered that some bacterial species have yet another version of GluRS instead of GlnRS. These organisms are found to have the D-GluRS and a GluRS enzyme (GluRS2) that only recognizes $tRNA^{Gln}$ and charges it with glutamate so that Gln-$tRNA^{Gln}$ can be formed using the indirect route (38). Because the different GluRSs have different specificities of tRNAs, we have limited the evolutionary study to the bacterial version of the D-GluRS.

The LeuRSs are also found to have the canonical pattern of evolution (25). In addition, the position of insertion of the editing domain in the catalytic domain varies between the bacterial version and the archaeal/eukaryal versions of the molecule that have the same position of insertion. The identity elements of $tRNA^{Leu}$ also vary among the 3 domains of life (31). Whereas the bacterial version of the molecule recognizes bases in the elbow region of $tRNA^{Leu}$, the archaeal version uses bases in the long variable arm of $tRNA^{Leu}$ to identify $tRNA^{Leu}$ from other tRNA molecules (39). The *H. sapiens* LeuRS uses a combination of bases in the elbow region and the long variable arm of $tRNA^{Leu}$ for the discrimination of $tRNA^{Leu}$. The discriminator base Ade73 is found to be a universal identity element for all $tRNA^{Leu}$ molecules. Because of the variable position of the insertion of the editing domain in the catalytic domain of LeuRS and the difference in identity elements of $tRNA^{Leu}$ in different domains of life, we limit the evolutionary study of LeuRS to the archaeal version of the molecule in this study.

Additional Results

Principal Component Analysis. Ten principal components (PCs) are required to account for ≈60% of the covariance in the motion during the GluRS:$tRNA^{Glu}$:Glu-AMP simulation (Fig. S3). Here, PC1 to PC3 couple the breathing motion at the CP1 region/acceptor stem interface with the breathing motion of the ACB domain of GluRS and the anticodon loop of $tRNA^{Glu}$ (Fig. S3). In higher PC modes, the coupling of the motion of the CP1 region with that of the ACB domain of GluRS remains, whereas the relative motion of the other regions of the tRNA increases. Even though the structures of the editing domain and the CTD that interact with the tRNA are different in LeuRS, a similar pattern is observed in the motion of the CTD that is coupled to the motion of the CP1 insertion, the Rossman fold, the 4HB, and parts of the editing domain (Fig. S4) in the simulation of the LeuRS:$tRNA^{Leu}$:Leu-AMP complex.

Characteristic Path Length Analysis. Removal of a node representing a tRNA nucleotide has a larger effect on the CPL of the dynamic protein:RNA network than the removal of a protein residue (Fig. S5). In the GluRS:$tRNA^{Glu}$:Glu-AMP simulation, the tRNA displays a larger amount of motion than the protein, and there are smaller correlation values of motion between nucleotides than between protein residues. In other words, the tRNA part of the network is sparser (average internucleotide degree = 2.58) and has larger edge distances (average internucleotide edge distance = 1.25) than the protein part (average interresidue degree = 7.02 and average interresidue edge distance = 0.46). Hence, the difference in CPL is affected to a larger degree when a nucleotide is removed than when a protein residue is removed.

The Z score of the change in CPL ($\Delta CPL_{k,\mathrm{rem}}$) upon node removal has been performed separately for the protein and the RNA (Fig. S6). Removal of nodes corresponding to either nucleotides in the elbow region of the tRNA that bind Mg^{2+} (U8 and G49) or nucleotides at the GluRS:tRNA interface (G26, the identity element U13, and U33) causes a significant change in CPL (Fig. S6*A*). Removal of nucleotides 26 and 44 also changes CPL drastically because they are important for communication between the anticodon stem and the GG arm. The identity element U13 and the anticodon show a small signal in this analysis. In the GluRS, a number of conserved residues near the adenylate-binding region (Ile 6, Pro 8, His 15 from the HIGH motif, Tyr 20, Arg 39, Glu 41, Ile 56, Gln 82, and Pro 234), the protein:tRNA interface (Val 145, Asp 160, Lys 180, Thr 186, and Arg 358), and interdomain interfaces on the protein (Arg 358, Trp 62, Phe 261, and Pro 263) (Z value ≥1 in Fig. S6*B*) are found to be significant in the CPL analysis. The CPL analysis predicts that a number of nucleotides in the tRNA play a large role in the communication pathways of the aaRS:tRNA complex. This effect is observed even in the CPL analysis of the GluRS:$tRNA^{Glu}$ crystal structure even though the particular residues that have a significant effect on the CPL are not exactly the same but are close to those in the dynamic protein:RNA network.

Regions close to the periphery of the aaRS:tRNA graph such as the nucleotides in the common arm and the residues in the ACB domain are found to have negative Z values in the CPL analysis. Removing these nodes reduces the CPL as one of the most distant nodes in the network is removed. The CPL analysis finds most of the significant regions that are important for allostery, but it also finds a number of residues and nucleotides that are not conserved (Table S2). Finally, the significance of some of the residues and nucleotides identified in this analysis can be understood in terms of the communication between local regions.

Community Analysis of Crystal Structure. The community analysis of the unweighted graph based on the crystal structure of the GluRS:$tRNA^{Glu}$:Glu-AMP complex has been performed. The number of communities increases from 13 in the correlation-based network to 14 for the static network with a community repartition difference of 0.44. Also, the nodes identified as critical for allostery correspond to nucleotides and amino acids with markedly lower conservation (Table S2). In general, the regions identified as critical for allostery were similar to the regions identified using the CPL analysis and the community analysis of the dynamic protein:RNA network. The residues and nucleotides important for allostery that have highest conservation are predicted using the community analysis of the dynamic protein:RNA network. The use of multiple X-ray structures or a simplified energy function to weight the network may capture more dynamic effects, but these procedures have not been tested.

Community Repartition Difference. ***Modifications to the tRNA.*** The community partitions for modified networks are compared with the community distribution for the wild-type GluRS:$tRNA^{Glu}$:Glu-AMP network. In each of the modified networks, the correlations between an interface nucleotide and its contacts in GluRS are scaled by a factor of one-half (Fig. S8). This results in weaker contacts in the interface of the modified protein:RNA networks, and the corresponding edges are represented by larger distances in the network.

The community distribution for each modified network is compared to the wild-type community distribution for the GluRS:$tRNA^{Glu}$:Glu-AMP network using the community repar-

tition difference. A random partition of 545 nodes (number of monomers) into a similar number of communities (11 to 15) because the wild-type network has a community repartition difference in the range [0.916–0.939]. The community distributions of all of the modified networks differ from the wild-type GluRS:tRNA^{Glu}:Glu-AMP community distribution by <0.5, indicating that the changes caused by these modifications are much smaller. In addition, most of the modifications that occur at or close to the identity elements (U13, the anticodon U35 and C36, and C4) or the active site (A76) have the largest effect on the community distribution. These nucleotides each make many contacts to the GluRS and so their mutations affect large numbers of shortest paths that are used to determine the communities.

The weakening of the interface edges of U13 causes some of the largest changes in the community distribution of the modified network. As measured by the community repartition difference, 65% of the node pairs remain grouped together in the same community, whereas 35% of the node pairs are split into separate communities. These edges are removed early in the Girvan–Newman cluster algorithm and, hence, have the largest overall effect on the community assignment of the nodes.

Modifications to the protein. To compare the results from the community distribution to alanine scan experiments on the protein, the networks were modified for each residue such that its contacts with the tRNA were weakened (Fig. S9). The residues that cause the largest difference in the community distribution for the modified networks were: residues near the active site, i.e., were interacting with the adenylate Glu-AMP or CCA hairpin of the tRNA (Gly 17, Ile 21, Asp 44, Arg 47, Gly 121, Val 177, Lys 180, Thr 186, Tyr 187), the residues that were interacting with the anticodon (Pro 445), or residues close to identity elements on the GG arm of the tRNA (Glu 282, Lys 309, Trp 312). This shows that most of the residues causing the largest change in community partition are located near important regions in the protein:RNA complex.

Modification to tRNA^{Leu}. The community partitions for modified networks are compared with the community distribution for the wild-type LeuRS:tRNA^{Leu}:Leu-AMP network. In each of the modified networks, the correlations between an interface nucleotide and its contacts in LeuRS are scaled by a factor of one-half (Fig. S10). Similar to Fig. S8, the nucleotides in the acceptor stem and the CCA hairpin significantly affect the community distribution of the wild-type LeuRS:tRNA^{Leu}:Leu-AMP network. In tRNA^{Glu}, the nucleotides on the GG arm helix (U11 and U13) and the anticodon loop (U35 and C36) affect the community distribution of the GluRS:tRNA^{Glu}:Glu-AMP network significantly. In contrast tRNA^{Leu} forms strong contacts between the CTD and both the loop on the GG arm (G19) and the variable arm (G47 and G47E) of tRNA^{Leu}. Hence, these nucleotides have a large effect on the community distribution of the LeuRS:tRNA^{Leu}:Leu-AMP network.

Parameter Sensitivity. In the network, edges connect pairs of nodes if the corresponding monomers are in contact, and 2 nonconsecutive monomers are said to be in contact if any heavy atoms (nonhydrogen) from the 2 monomers are within 4.5 Å of each other in at least 75% of the frames analyzed. The community for the wild-type GluRS:tRNA^{Glu}:Glu-AMP network is compared to that for the GluRS:tRNA^{Glu}:Glu-AMP network made using different parameters for the contact distance (3.5–5 Å) and the percentage of frames (65% to 85%) to be considered in contact in Fig. S11. When the contact distance is $\leq$4.0 Å, the number of edges for each node falls down drastically. This results in a number of isolated (no edges), single-node communities (e.g. maximum of 100 communities when distance cutoff = 3.5 Å using 85% of the frames in the simulation) and high community repartition differences. At cutoff distances from 4.5 to 5.0 Å, the community partition difference is in the range of 0.25–0.48. This shows that the effect of the different parameters on community distribution is minor as compared with random partitions of the nodes into communities for cutoff distance $\geq$4.5 Å

1. Sekine S, et al. (2003) ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J* 22:676–688.
2. Humphrey W, Dalke A, Schulten K (1996) VMD—Visual molecular dynamics. *J Mol Graphics* 14:33–38.
3. Perona JJ, Rould MA, Steitz TA (1993) Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemistry* 32:8758–8771.
4. Fukunaga R, Yokoyama S (2005) Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation. *J Mol Biol* 346:57–71.
5. Eswar N, Eramian D, Webb B, Shen M, Sali A (2008) Protein structure modeling with MODELLER. *Methods Mol Biol* 426:145–159.
6. Cusack S, Yaremchuk A, Tukalo M (2000) The 2 A crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J* 19:2351–2361.
7. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comp Chem* 26:1781–1802.
8. Foloppe N, MacKerrell AD, Jr (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comp Chem* 21:86–104.
9. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J Mol Graphics* 8:52–56.
10. O'Donoghue P, Sethi A, Woese CR, Luthey-Schulten ZA (2005) The evolutionary history of Cys-tRNA^{Cys} formation. *Proc Natl Acad Sci USA* 102:19003–19008.
11. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
12. Eargle J, Black AA, Sethi A, Trabuco LG, Luthey-Schulten Z (2008) Dynamics of recognition between tRNA and elongation factor Tu. *J Mol Biol* 377:1382–1405.
13. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An *N*·log(*N*) method for Ewald sums in large systems. *J Chem Phys* 89:10089–10092.
14. Auffinger P, Westhof E (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J Mol Biol* 269:326–341.
15. Ichiye T, Karplus M (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct Funct Genet* 11:205–217.
16. Hunenberger PH, Mark AE, van Gunsteren WF (1995) Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol* 252:492–503.
17. Young MA, Gongloni S, Superti-Furga G, Roux B, Kuriyan J (2001) Dynamic coupling between the sh2 and sh3 domains of c-src and hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 105:115–126.
18. Tai K, Shen T, Börjesson U, Philipopoulos M, McCammon JA (2001) Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys J* 81:715–724.
19. Glykos NM (2006) Software news and updates. CARMA: A molecular dynamics analysis program. *J Comput Chem* 27:1765–1768.
20. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Syst Biol* 2:1–12.
21. Knox EG (1964) The detection of space-time interactions. *Appl Stat* 13:25–30.
22. Rand W (1971) Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 66:846–850.
23. O'Donoghue P, Luthey-Schulten Z (2003) On the evolution of structure in the aminocyl-tRNA synthetases. *Microbiol Mol Bio Rev* 67:550–573.
24. Sethi A, O'Donoghue P, Luthey-Schulten Z (2005) Evolutionary profiles from the QR factorization of multiple sequence alignments. *Proc Natl Acad Sci USA* 102:4045–4050.
25. Woese CR, Olsen GJ, Ibba M, Soll D (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* 64:202–236.
26. McGinnis S, Madden TL (2004) BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32:W20–W25.
27. Wheeler DL, et al. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 28:10–14.
28. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
29. Lee J, Hendrickson T (2004) Divergent anticodon recognition in contrasting glutamyl-tRNA synthetases. *J Mol Biol* 344:1167–1174.
30. Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7:382.
31. Ibba M, Francklyn C, Cusack S, eds (2005) *The Aminoacyl-tRNA Synthetases* (Landes Bioscience, Georgetown, TX).
32. Sekine S, et al. (1996) Major identity determinants in the ''augmented D helix'' of tRNA^{Glu} from *Escherichia coli*. *J Mol Biol* 256:685–700.
33. Sekine S, Nureki O, Tateno M, Yokoyama S (1999) The identity determinants required for the discrimination between tRNA^{Glu} and tRNA^{Asp} by glutamyl-tRNA synthetase from *Escherichia coli*. *Eur J Biochem* 261:354–360.

34. Madore E, et al. (1999) Effect of modified nucleotides on *Escherichia coli* $tRNA^{Glu}$ structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. *Eur J Biochem* 266:1128–1135.
35. Sekine S, Nureki O, Shimada A, Vassylyev D, Yokoyama S (2001) Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. *Nat Struct Biol* 8:203–206.
36. Rould M, Perona J, Söll D, Steitz T (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science* 246:1135–1142.
37. Wilcox M, Nirenberg M (1968) Transfer RNA as a cofactor coupling amino acid synthesis with that of a protein. *Proc Natl Acad Sci USA* 61:229–236.
38. Skouloubris S, Ribas de Pouplana L, De Reuse H, Hendrickson T (2003) A noncognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proc Natl Acad Sci USA* 100:11297–11302.
39. Soma A, Uchiyama K, Sakamoto T, Maeda M, Himeno H (1999) Unique recognition style of tRNA(Leu) by *Haloferax volcanii* leucyl-tRNA synthetase. *J Mol Biol* 293:1029–1038.

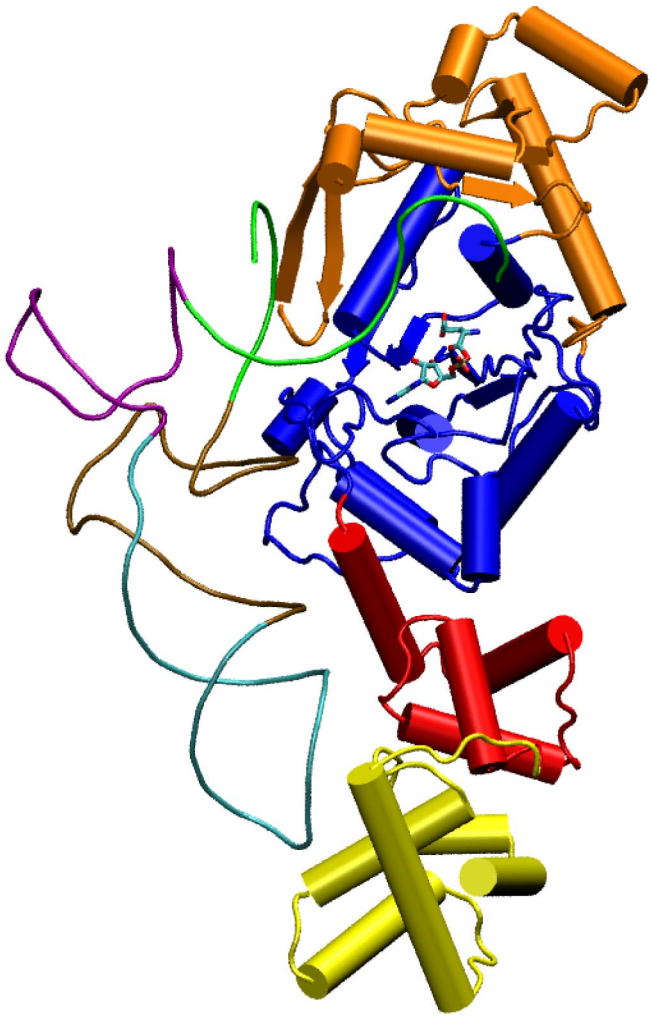

Fig. S1. The structure of the GluRS:tRNA^{Glu}:Glu-AMP complex. The residues and nucleotides in the protein and RNA are colored by the domain they belong to: Rossman fold (blue), CP1 insertion (orange), 4HJ (red), ACB domain (yellow), acceptor arm (green), GG arm (tan), anticodon arm (cyan), and common arm (purple). The adenylate Glu- AMP is shown in licorice colored by atom type.

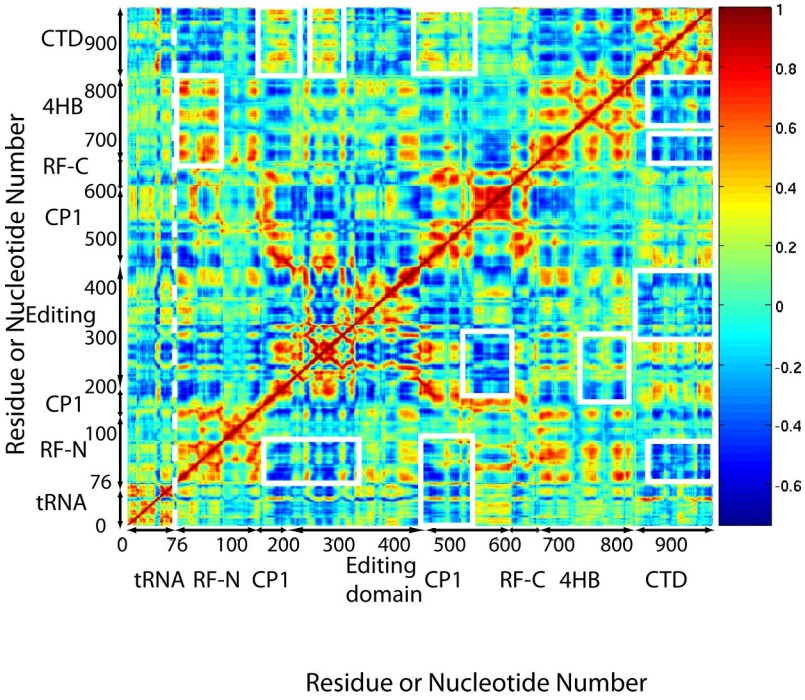

Fig. S2. Correlation analysis of the motion during 16 ns of the equilibration of the LeuRS:tRNA^{Leu}:Leu-AMP complex. The monomers with highly correlated motion are orange or red, whereas regions that are anticorrelated are shown in blue. Distant regions that are correlated are marked in white rectangles above the diagonal, whereas anticorrelated regions are marked below the diagonal.

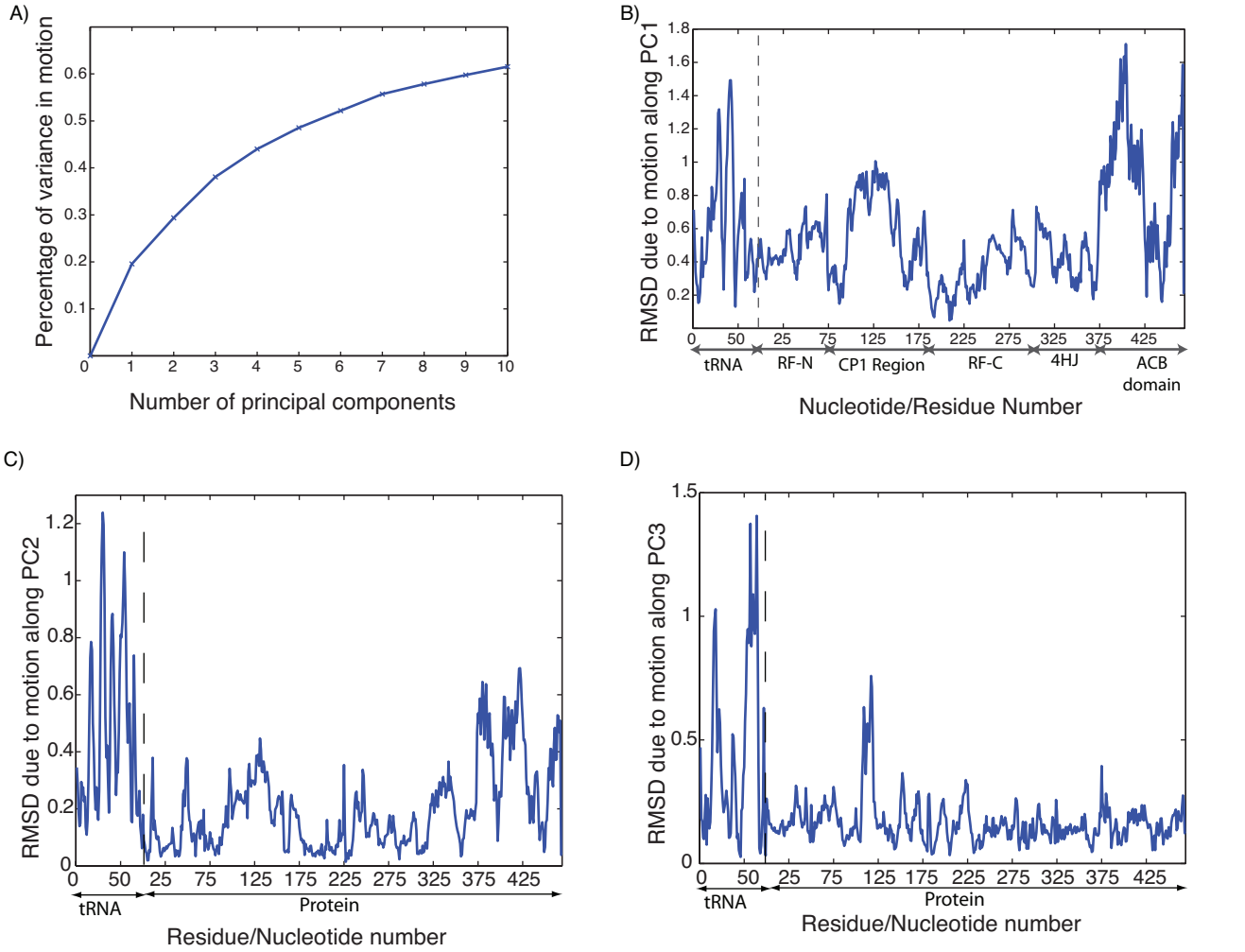

Fig. S3. Principal component analysis of the motion during the 16-ns equilibration of the GluRS:tRNA^{Glu}:Glu-AMP complex. (*A*) Percentage of variance in motion along the first 10 principal components is shown. (*B–D*) The RMSD of each monomer in the complex due to motion along the first (*B*), second (*C*), and third (*D*) principal component modes is shown.

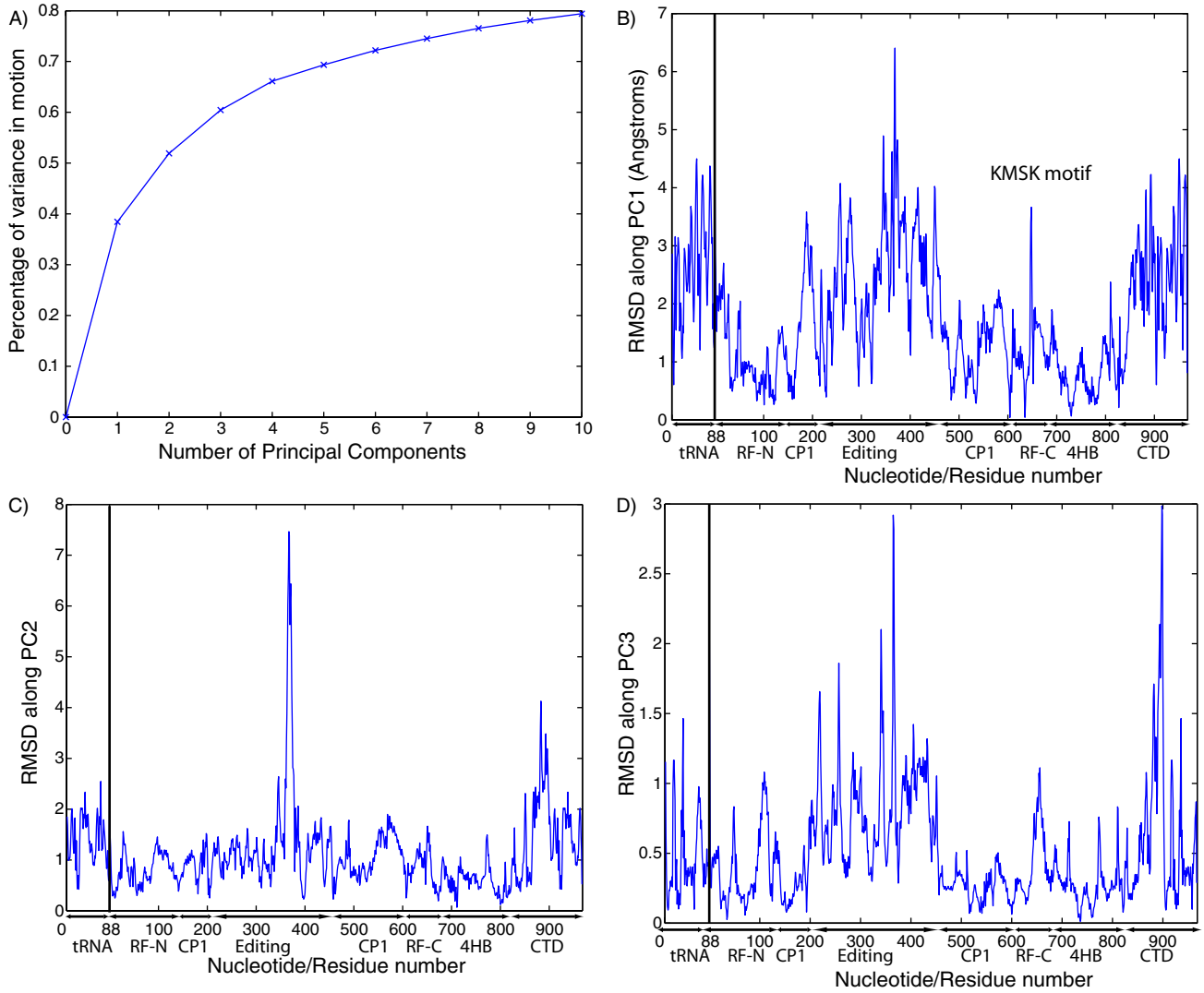

Fig. S4. Principal component analysis of the motion during the 16-ns equilibration of the LeuRS:tRNA^{Leu}:Leu-AMP complex. (*A*) Percentage of variance in motion explained by the first 10 principal components are shown. (*B–D*) The RMSD of each monomer in the complex due to motion along the first (*B*), second (*C*), and third (*D*) principal component modes is shown.

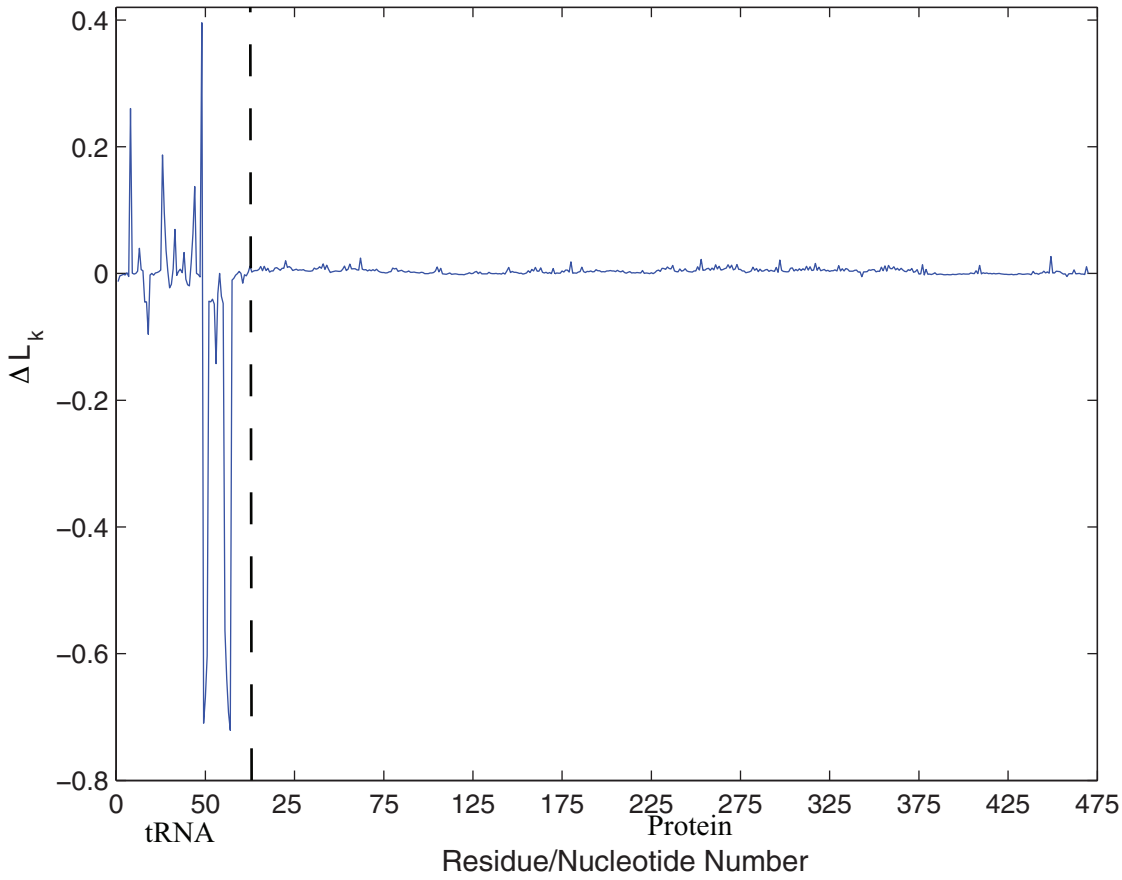

Fig. S5. Characteristic path length of the dynamic protein:RNA network made from the GluRS:tRNA^{Glu}:Glu-AMP simulation. The RNA nucleotides have a larger effect on the CPL than the protein residues.

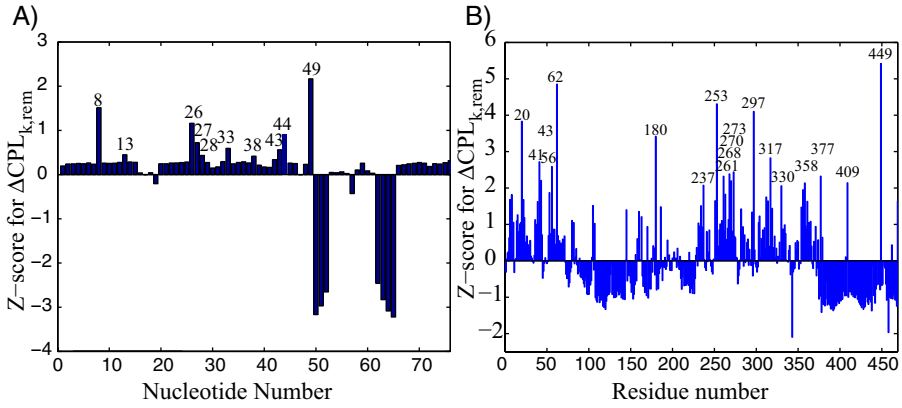

Fig. S6. Difference in characteristic path length: The *Z* value for the difference in CPL is shown for node removal in $tRNA^{Glu}$ (*A*) and GluRS (*B*) in the GluRS:$tRNA^{Glu}$:Glu-AMP network. The monomers that are significant in the *Z* value are marked in *A* (*Z* value ≥0.4) and *B* (*Z* value ≥2).

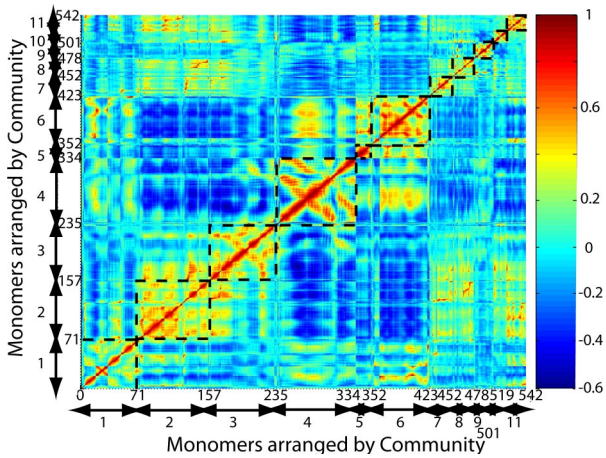

Fig. S7. Correlation between communities. The correlation of the residues and nucleotides in the GluRS pretransfer complex is shown. The correlation of the monomers shown in Fig. 1 has been rearranged by the community they belong to.

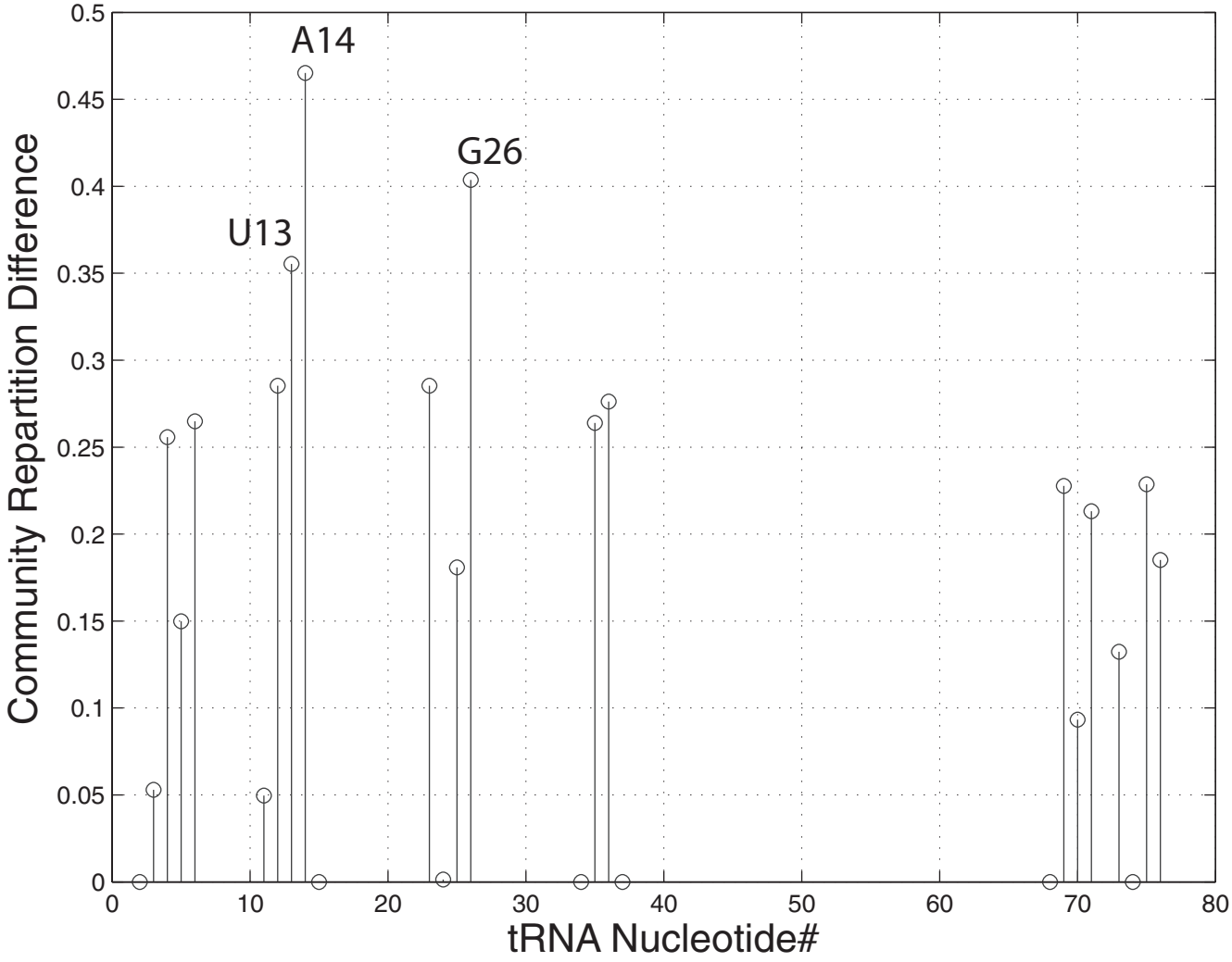

Fig. S8. Community repartition differences for nucleotide mutations at the GluRS:tRNA interface. Correlation values for each mutation have been reduced by 50%. Nucleotide mutations that resulted in a community repartition difference $>$0.3 are labeled.

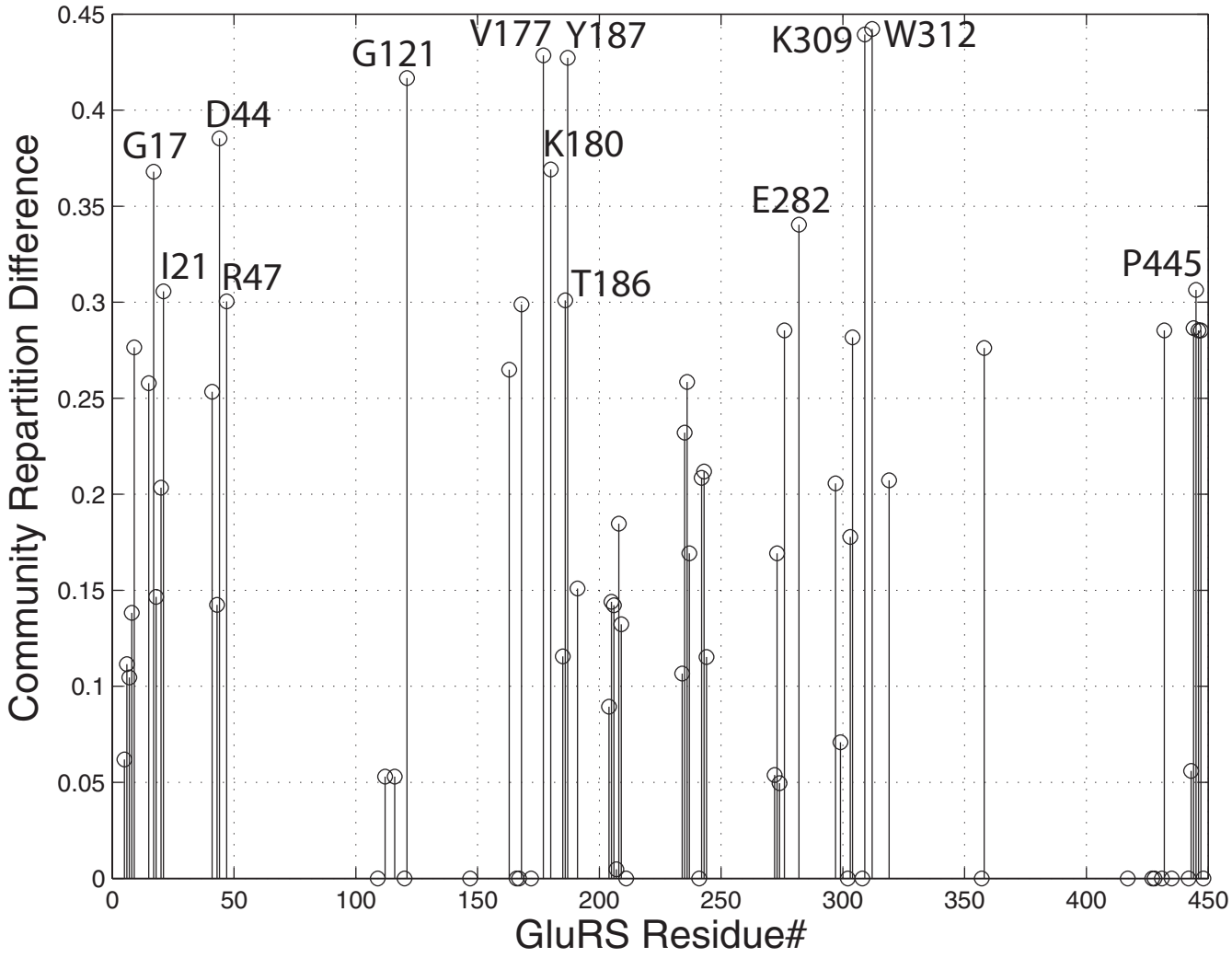

Fig. S9. Community repartition differences for amino acid mutations at the GluRS:tRNA interface. Correlation values for each mutation were reduced by 50%. Amino acid mutations that resulted in a community repartition difference >0.3 are labeled.

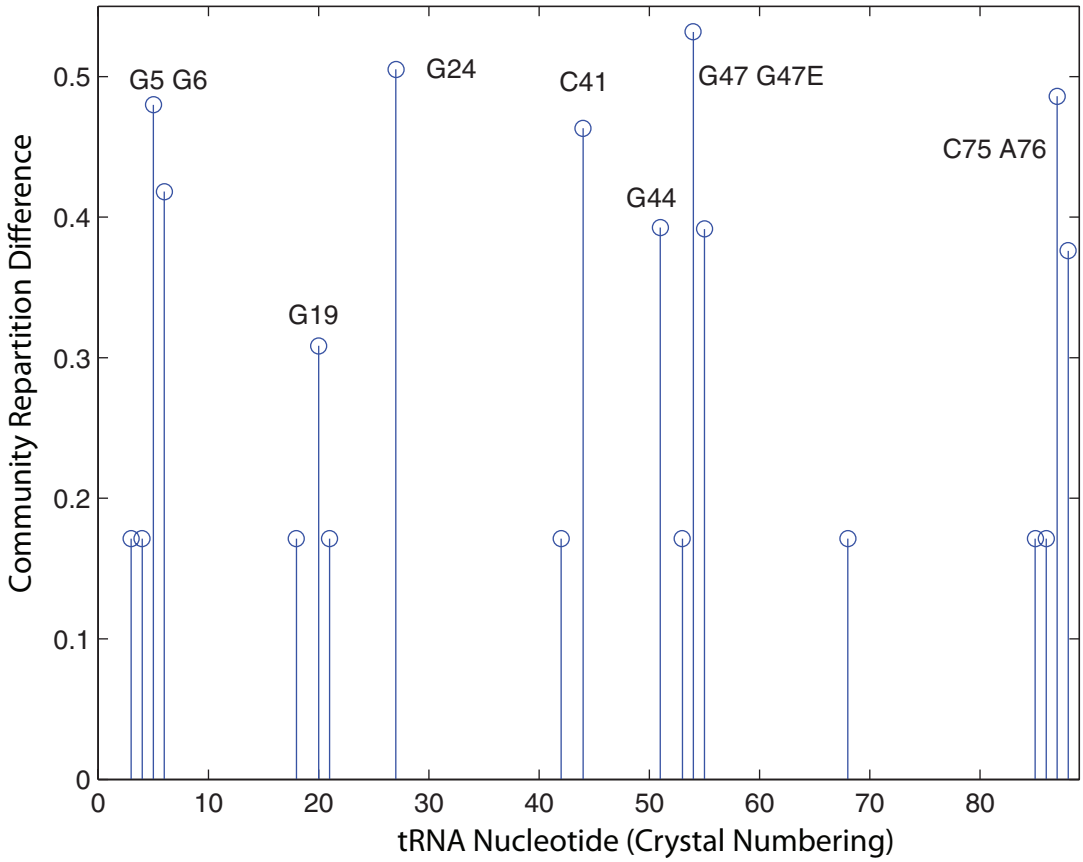

Fig. S10. Community repartition differences for nucleotide mutations at the LeuRS:tRNA interface. Correlation values for each mutation have been reduced by 50%. Nucleotide mutations that resulted in a community repartition difference $>$0.3 are labeled.

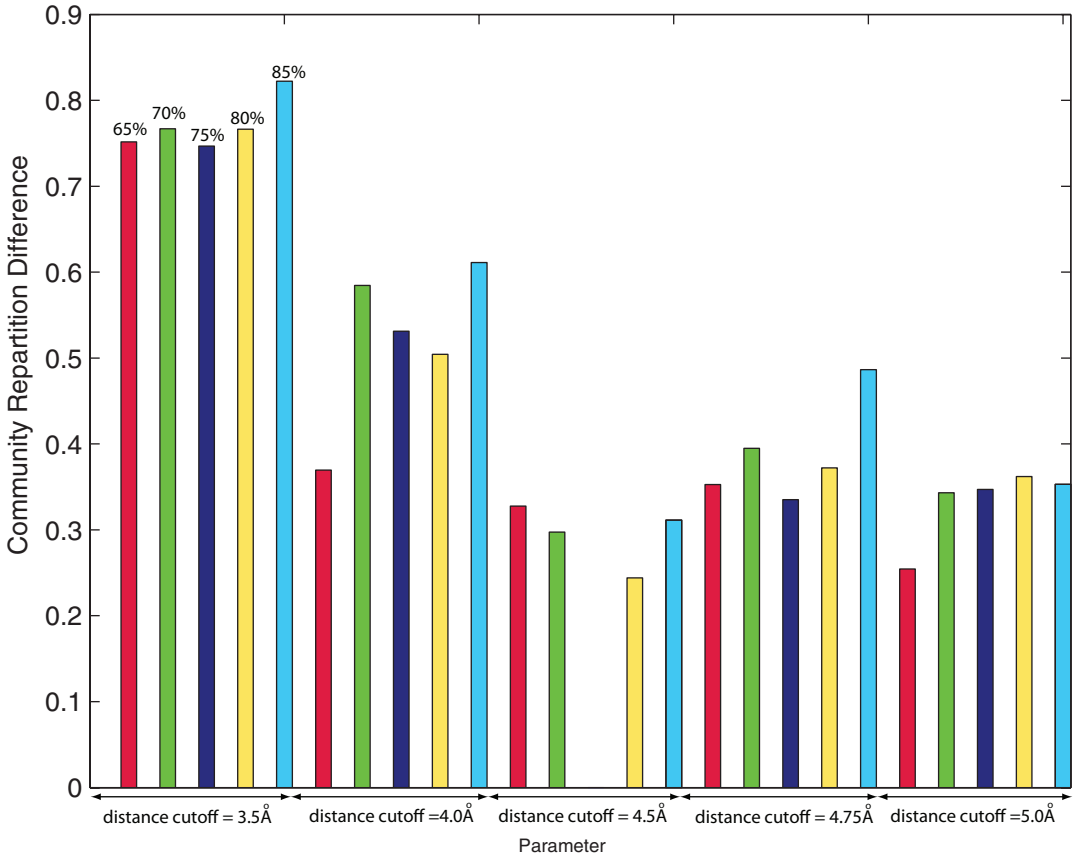

Fig. S11. Community partition comparison for the network representing the GluRS:tRNA^{Glu}:Glu-AMP complex for different parameters in defining the network. The contacts that form edges in the network are decided based on 2 parameters: (*i*) distance cutoff for defining a contact in a given frame of the simulation and (*ii*) the percentage of frames that the contact has to be formed in the simulation.

Table S1. The shortest distance of each identity element ($D^0_{i,A76}$) and number of suboptimal paths (N_{sop}) of LeuRS from A76 in the LeuRS:tRNA^{Leu}:Leu-AMP network

Source	$D^0_{i,A76}$	N_{sop}
C20A	5.41	281
A46	4.01	720
G47	3.73	702
U47H	10.19	276
A73	1.26	78

Table S2. Comparison of the nucleotides predicted to be crucial for allostery in the GluRS:tRNA^{Glu}:Glu-AMP system based on the CPL (Node), CPL (Edge), and community analyses and their conservation

Monomer	Conservation	CPL (Node)	CPL (Edge)	Community	CPL (ContactMap)	Community (ContactMap)
C4	74.1	No	Yes	No	Yes	No
C5	85.2	No	Yes	No	No	No
C6	48.1	No	Yes	No	Yes	Yes
U8	100	Yes	No	Yes	No	Yes
C9	57.1	No	No	No	No	Yes
U11	100	No	No	Yes	Yes	No
C12	77.8	No	Yes	No	Yes	Yes
U13	66.7	Yes	Yes	No	Yes	No
A14	100	No	Yes	Yes	No	No
G15	66.7	No	Yes	No	Yes	Yes
G19	100	No	No	No	No	Yes
G23	64.4	No	No	No	Yes	No
A24	82.2 G	No	No	No	Yes	No
C25	100	No	Yes	Yes	No	No
G26	41.7	Yes	Yes	No	Yes	No
C27	52.1	Yes	No	No	No	No
G28	56.3	Yes	No	No	No	No
C31	71.4	No	No	No	No	Yes
U33	100	Yes	No	Yes	Yes	No
U35	100	No	Yes	No	Yes	No
C36	100	No	Yes	No	Yes	No
A38	29.2	Yes	No	Yes	Yes	Yes
G39	66.7	No	No	No	Yes	Yes
G43	68.8	Yes	No	No	No	No
A44	47.9	Yes	No	No	No	No
A46	76.2	No	No	No	No	Yes
G49	22.9	Yes	No	Yes	No	Yes
G50	80.4	No	No	No	No	Yes
U59	30.4	No	No	No	No	Yes
C65	15.2	No	No	No	No	Yes
G67	2.17	No	No	No	No	Yes
G68	88.9	No	Yes	No	No	No
G69	70.4	No	Yes	No	Yes	No
G70	74.1	No	Yes	No	No	No
U71	24.4	No	No	No	Yes	Yes
A73	73.9	No	No	No	Yes	No
C74	100	No	No	No	Yes	No
C75	100	No	Yes	No	No	No
A76	100	No	Yes	No	Yes	Yes
Arg5	97.2	No	No	No	No	Yes
Ile6	100 hp	Yes	No	No	No	Yes
Pro8	100	Yes	No	Yes	No	Yes
Asp13	16.7	No	No	No	No	Yes
His15	100	Yes	No	No	No	Yes
Tyr20	100 Tyr/Arg	Yes	No	Yes	No	No
Ile21	14.3	Yes	No	No	No	No
Ala22	85.7	Yes	No	No	Yes	No
Phe36	83.3	No	No	No	No	Yes
Arg39	100	Yes	No	No	No	Yes
Ile40	97.2 hp	No	No	No	No	Yes
Glu41	100 Asp/Glu	Yes	No	No	No	No
Thr43	100	Yes	No	No	No	No
Glu53	51.4	Yes	No	No	No	No
Arg55	5.6	No	No	No	No	Yes
Ile56	100 hp	Yes	No	Yes	Yes	Yes
Trp62	89	Yes	No	Yes	Yes	No
Tyr80	77.1	Yes	No	No	No	No
Gln82	100	Yes	No	Yes	No	Yes
Tyr89	89	No	No	Yes	No	No
Tyr103	89	No	No	Yes	No	No
Ala105	14.3	Yes	No	No	No	No
Phe106	97 aro	No	No	Yes	No	No
Glu107	14.3 Asp/Glu	Yes	No	No	Yes	No
Arg116	26.7	No	No	No	Yes	No
His144	11.1	No	No	No	No	Yes
Val145	75.6	Yes	No	No	No	No

Monomer	Conservation	CPL (Node)	CPL (Edge)	Community	CPL (ContactMap)	Community (ContactMap)
Asp160	100	Yes	No	No	No	No
Arg163	62.2	Yes	Yes	No	Yes	No
Val166	42.2	No	Yes	No	No	No
Tyr168	44.4 Phe/Tyr	No	Yes	No	No	No
Glu172	26.7	No	Yes	No	Yes	No
Ile173	27.8	No	No	No	No	Yes
Asp175	94.4	No	No	No	No	Yes
Val177	95.6 hp	No	Yes	No	No	No
Lys180	100 Lys/Arg	Yes	Yes	No	Yes	Yes
Thr186	75.5	Yes	Yes	No	Yes	No
His232	100	No	No	Yes	No	No
Pro234	91.1	No	No	Yes	No	No
Arg237	15.6	No	Yes	No	Yes	No
Lys241	31.1 Lys/Arg	No	Yes	No	Yes	Yes
Lys243	95	No	No	No	Yes	Yes
Ile244	100 hp	No	No	No	No	Yes
Thr251	22.2	Yes	No	No	No	No
Leu253	100hp	Yes	No	No	Yes	Yes
Asp254	19.4	No	No	No	No	Yes
Lys257	11.1	No	No	No	No	Yes
Phe261	94 Phe/Tyr	Yes	No	Yes	No	No
Pro263	88.9	Yes	No	No	No	No
Arg267	11.1	No	No	No	No	Yes
Asn268	97.5	Yes	No	No	No	No
Leu270	100 hp	Yes	No	No	No	No
Met273	100 hp	Yes	Yes	No	Yes	No
Gly274	91	No	No	Yes	No	No
Ser276	66.7 Ser/Thr	No	Yes	No	No	No
Gln281	13.3	No	Yes	No	No	No
Glu282	97.5 Asp/Glu	Yes	No	No	Yes	No
Trp295	5	Yes	No	No	No	No
Arg297	45.6	Yes	Yes	No	Yes	No
Pro303	13.3	No	Yes	No	No	No
Val304	44.4	Yes	No	No	Yes	No
Phe305	89	No	No	Yes	No	Yes
Asp306	100 Asp/Asn	No	No	Yes	No	No
Lys309	95	No	No	No	Yes	No
Leu310	937	No	No	Yes	No	No
Trp312	46.6	Yes	Yes	Yes	Yes	No
Asn314	97 Asn/Gln	Yes	No	No	Yes	No
Tyr317	66.7	Yes	No	No	Yes	Yes
Arg319	70 Lys/Arg	Yes	No	No	Yes	No
Arg330	17.8	Yes	No	No	Yes	No
Phe334	22.2	No	No	No	No	Yes
Arg354	55.6	Yes	No	No	No	No
Arg356	8.9	Yes	No	No	No	No
Arg358	100 Lys/Arg	Yes	Yes	No	Yes	Yes
Leu359	80 hp	No	No	Yes	No	No
Asp360	5.6	No	No	No	No	Yes
Thr361	35.6	Yes	No	No	No	No
Lys368	11.1	No	No	No	No	Yes
Tyr371	38.9 aro	No	No	No	No	Yes
Tyr377	35.6	Yes	No	No	No	No
Glu409	7	Yes	No	No	Yes	No
Arg417	10	No	No	No	Yes	No
Gln432	15.6	No	Yes	No	Yes	No
Glu443	2.2	No	Yes	No	Yes	No
Thr444	100	No	Yes	No	No	No
Pro445	80	No	Yes	No	No	No
Glu449	66.7 Asp/Glu	Yes	No	Yes	No	No
Average conservation		67.6	65.8	85.8	63.4	60.9

Contact Map refers to applying the CPL (Node) and Community analyses to the unweighted contact map network from the crystal structure of the adenylate complex (PDB ID: 1N78) with adenylate analog modified to Glu-AMP. ″hp″ refers to hydrophobic residues, whereas ″aro″ refers to aromatic residues.

Table S3. The optimal community distribution for the GluRS:tRNA^{Glu}:Glu-AMP network

Community	tRNA nucleotide	Protein residue
C-1	C4, and C5-G68	Parts of CP1 insertion (85–98 and 155–168), strands 4 and 5, and helices 4 and 5 of the Rossman fold (187–205 and 209–225, 232, 300)
C-2	—	Glu-AMP, strands 1 to 3, helices 1 to 3 of Rossman fold (1–6,14–38, 54–80), and GluRS specific insertion (263–273 and 284–295)
C-3	A37	Four helix junction and connector region from Rossman fold to four helix junction (250–262, 307–376)
C-4	U33 to C36	Anticodon binding domain (residues 375 to 468)
C-5	A73	CP1 helix (residues 107 to 125)
C-6	CCA hairpin, G1- C72, and G2	Loops in RF (43–47) and part of CP1 insertion (97–106, 122–154, and 169–186)
C-7	C3, C12, and G69 to U71	Loops in RF (206–208, 234–249 and 301–306)
C-8	GG arm (8–9, U11, 13–15, 20–24, A46, C48)	Loops in RF (272–282 and 292–299)
C-9	Tertiary contacts between GG arm and common arm (6–7, 16–19 and 49–67)	—
C-10	Anticodon stem (G10, 25–32 and 38–45)	Lys309
C-11	—	Loops in RF (7–12, 39–42, 48–53, Pro71, Val73, and 79–84)
C-12	U20A	—
C-13	U59	—

Table S4. The monomers involved in each interaction predicted to be crucial for allostery in the GluRS:tRNA^{Glu}:Glu-AMP system and their conservation

Monomer	Conservation	Monomer	Conservation
Pro234 (1)	78%	Pro228 (2)	100%
Thr156 (1)	50%	Asn170 (6)	60%
Pro234 (1)	78%	His232 (7)	97%
Tyr20 (1)	100% Tyr/Arg	His232 (7)	97%
Tyr89 (1)	89%	Gln82 (11)	100%
Trp62 (2)	89%	Phe261 (3)	94% Phe/Tyr
Gly17(2)	97%	His232(7)	97%
Gly274 (2)	91%	A14 (8)	100%
Ile56 (2)	100% hydrophobic	Pro8 (11)	100%
Leu359 (3)	75% hydrophobic	Glu449 (4)	80% Asp/Glu
Leu310 (3)	93%	Asp306 (7)	100% Asp/Asn
Trp312 (3)	64% aromatic	C25 (10)	100%
U33(4)	100%	A38(10)	23%
Phe106(5)	97% Aromatic	Tyr103 (6)	89%
Glu41 (6)	100% D/E	Thr43 (11)	100%
Phe305(7)	89%	U11(8)	93%
U8(8)	100%	G49(9)	82%
U11(8)	93%	C25(10)	100%

The numbers within parentheses denote which community the monomer belongs to. In addition to the identity elements, these monomers are predicted to be important for allostery.

Table S5. The monomers that occur in the majority (>50%) of suboptimal paths from the source nucleotide (in column 1) to A76 are given in column 2

Source	Monomers
C34 (315)	Pro8 (255), Glu41 (274), Thr43 (278), Ile56 (244), Ala59 (235), Trp62 (239), Phe261 (189), Thr361 (171), Gly446 (194), and Glu449 (208)
U35 (196)	Pro8 (141), Glu41 (156), Thr43 (160), Ile56 (127), Ala59 (117), Trp62 (133), Phe261 (117), Met355 (146), Leu359 (146), Thr361 (151), and Pro445 (195)
C36 (204)	Pro8 (145), Glu41 (162), Thr43 (167), Ile56 (137), Ala59 (127), Trp62 (136), Phe261 (145), Asn314 (183), Leu359 (204), and Asp360 (166)
A37 (230)	Glu-AMP (158), Tyr20 (116), and Arg319 (124)
U11 (209)	C 25 (183) and Trp312 (16)
U13 (74)	Glu-AMP(47),Tyr20 (61), and Gly274 (60)

The numbers in parentheses represent the number of occurrences of the monomer in the suboptimal paths.

Table S6. The optimal community distribution for the LeuRS:tRNA^{Leu}:Leu-AMP network identified in Fig. 4

Community	tRNA nucleotide	Protein residue
1	73 to 76	Part of CP1 insertion (164–180 and 461–533), β -strands 4 and 5 and α -helix 4 of Rossman fold (607–640)
2	—	Part of N-terminus extension, β -strands 1 to 3, and α -helices and 2 of Rossman fold (22–43, 48–79, and 134–151), and LeuRS specific insertion (642–690)
3	—	N-terminus extension, 4-helix bundle and loop to C terminus domain (1–20, and 691–833)
4	A 20A and variable arm	Part of C-terminus domain interacting with variable arm (837–866 and 916–967)
5	—	Strands and one helix in editing domain (206–326, 332–353, and 431–446)
6	Tertiary contacts between GG arm and common arm	Part of C terminus domain interacting with these loops (867–915)
7	Acceptor stem, GG arm helix, common arm helix, and anticodon arm	-
8	—	α -helices in CP1 insertion that is LeuRS specific (152–165, 533–604, and 629–633)
9	—	α -helical insertion that follows β -strand 2 of Rossman fold (85–135)
10	—	Zn^{2+} binding motif of CP1 insertion (182–205, and 435–459)
11	—	Loop in editing domain (358–375)
12	G 1	Loops (231–234, 327–331, and 354–358), and α -helices in editing domain (376–430)

Ranges of protein residues that are part of a community are included within parentheses.

Table S7. The monomers involved in each interaction predicted to be crucial for allostery in the LeuRS:tRNA^{Leu}:Leu-AMP system and their conservation

Monomer	Conservation	Monomer	Conservation
Gly642 (1)	100%	Ala681 (2)	100% Gly/Ala/Ser
Phe466 (1)	92.3%	His181 (6A)	81% His/Tyr
Lys169 (1)	84.6% Lys/Arg/Gln	Asp565 (6B)	88.5% Asp/Glu/Asn
Tyr44 (1)	96.2%	Ser46 (6C)	96.2% Ser/Asn
Pro489 (1)	100%	G6 (G906) (8)	80.6%
Arg750 (2)	92.3% Lys/Arg	Ala681 (3)	100% Gly/Ala/Ser
Trp601 (2)	100% aromatic	Phe598 (6B)	92.3%
Trp601 (2)	100% aromatic	Glu597 (6B)	80.7% Asp/Glu
His81 (2)	80.7%	Tyr44 (6C)	96.2%
His48 (2)	100%	Ser654 (6C)	100%
Glu836 (3)		Arg840 (4A)	Backbone contact variable
Tyr767 (3)	84.6%	C41 (C944) (8)	80.6%
Lys870 (4A)		Ala874 (4B)	Variable backbone contact
G19 (G920) (4B)	100%	C20A (C922) (8)	91.7%
His181 (6A)	81% His/Tyr	Ile458 (12A)	100% hydrophobic
Asp197 (6A)	92.3%	Ala91 (9)	92.3% hydrophobic
Thr152 (6B)	100% Thr/Ser	Trp80 (6C)	100% aromatic
Phe312 (12A)	100% Phe/Tyr	Leu393 (12B)	81%
Ala375 (12B)	65.4%	Glu377 (12C)	69.3% Asp/Glu

The numbers within parentheses denote which community the monomer belongs to. In addition to the identity elements, these monomers are predicted to be important for allostery.

Table S8. List of organisms in the QR set for D-GluRS and tRNA^{Glu}

Species	D-GluRS	tRNA
Thermus thermophilus	X	X
Escherichia coli		X
Desulfotalea psychrophila	X	
Arcobacter butzleri	X	
Candidatus Blochmannia floridanus	X	
Buchnera aphidicola	X	
Vibrio cholerae	X	
Methylobacterum sp.	X	
Orientia tsutsugamushi Boryong	X	
Rickettsia sibirica	X	
Wolbachia endosymbiont of Drosophila ananassae	X	
Anaplasma marginale	X	
Petrotoga mobilis	X	
Thermotoga maritima	X	
Thermotoga petrophila		X
Tropheryma whipplei	X	
Corynebacterium efficiens	X	X
Opitutaceae bacterium	X	
Frankia sp.	X	
Blastopirellula marina	X	
Rhodopirellula baltica	X	
Gloeobacter violaceus	X	
Deinococcus radiodurans	X	
Propionibacterium acnes	X	X
Chlamydophila pneumoniae	X	
Psychrobacter arcticus	X	
Oceanobacter sp.	X	
Rhizobium etli	X	
Plesiocystis pacifica	X	
Treponema pallidum	X	X
Borrelia burgdorferi	X	X
Vicitvallis vadensis	X	
Alkaliphilus metallredigens	X	
Chloroflexus aggregans	X	
Herpetosiphon aurantiacus	X	
Rubrobacter xylanophilus	X	
Microscilla marina	X	
Pedobacter sp.	X	
Flavobacter bacterium	X	
Salinibacter ruber	X	
Chlorobium phaeobacteroides	X	
Rhodopirellula baltica		X
Lactobacillus johnsonii		X
Lactobacillus brevis		X
Candidatus Ruthia magnifica str.		X
Candidatus Pelagibacter ubique		X
Oenococcus oeni		X
Neisseria gonorrhoeae		X
Ehrlichia canis str. Jake		X

Table S9. List of organisms in the QR set for archaeal LeuRS and tRNA^{Leu}

Species	LeuRS	tRNA
Pyrobaculum aerophilum	X	
Sulfolobus solfataricus	X	X
Sulfolobus tokodaii	X	X
Sulfolobus acidocaldarius	X	
Sulfolobus solfataricus	X	X
Aeropyrum pernix	X	
Methanocaldococcus jannaschii	X	
Methanococuss maripaludis	X	
Methanosarcina mazei	X	
Methanococcoides burtonii	X	
Archaeoglobus fulgidus	X	X
Thermococcus kodakarensis	X	X
Pyrococcus horikoshii	X	
Methanothermobacter thermautotrophicus	X	
Methanosphaera stadtmanae	X	
Methanopyrus kandleri	X	X
Methanospirillum hungateii	X	X
Nanoarchaeum equitans	X	X
Thermoplasma volcanium	X	
Thermoplasma acidophilum	X	X
Picrophilus torridus	X	
Ferroplasma acidarmanus	X	
Natromonas pharaonis	X	
Haloarcula marismortuii	X	
Haloferax volcanii		X
Methanoculleus marisnigri		X
Pyrobaculum calidifontis		X
Pyrococcus abyssi		X
Methanosarcina acetivorans		X
Halobacterium sp.		X
Methanosarcina barkeri		X
Hyperthermus butylicus		X
Methanosaeta thermophila		X
Picrophilus torridus		X
Methanocorpusculum labreanum		X